Finding relations between entities in a knowledge graph: Case artists of the Getty Union List of Artist Names (ULAN)

Heikki Rantala[1], Eero Hyvönen[1,2], and Petri Leskinen[2,1]

[1] Semantic Computing Research Group (SeCo), Aalto University, Finland
[2] Helsinki Centre for Digital Humanities (HELDIG), University of Helsinki, Finland

Abstract. This poster paper presents work on developing a knowledge-based approach for searching semantically interesting relations between two entities, or groups of them, in a knowledge graph using faceted search. We demonstrate the new search method by searching relations between artists in the Getty ULAN knowledge graph.

1 Knowledge based approach to relational search

In relational search (RS) [10,8,4,3,11] the goal is to find new relations in a knowledge graph (KG). A key challenge in RS is how to separate "interesting" connections from not interesting ones, and how to explain the connections to the end-user. In [6] a knowledge-based method to address these issues was presented, and applied to finding relations between people and places. The general idea is to transform a RS problem into a faceted search problem and then use ready-to-use-tools, such as Sampo-UI [7], for problem solving. In this paper, this method is applied as a case study to the Getty ULAN[3] [2] KG which includes rich data about over 300 000 internationally important artists.

In our approach we use SPARQL CONSTRUCT queries to precalculate instances of the class Relation between two entities, in this case two artists, based on predefined forms that represent interesting connection types in the data using expert domain knowledge. These connections and their explanations can then be explored efficiently using faceted search. The facets represent classes of the endpoints of the connection, such as two artists, or their properties, such as nationality. The facets can also represent the properties of the connection itself, such as the type of the connection or its relative importance. Predescribing the types of interesting connections means that uninteresting connections between people, like both being instances of the owl:Thing class, are eliminated. However, this method requires some KG specific knowledge engineering work. Furthermore, some unanticipated truly serendipitous connections may be missed.

Below is an example of a CONSTRUCT QUERY used to create Relation instances. The query finds two people who are connected by having the same

[3] https://www.getty.edu/research/tools/vocabularies/ulan/index.html

teacher, and creates instances of the Relation class that have those two people as the endpoints of the directed connection: the relationSubject and the relationObject. It also creates a human readable explanation of the relation as the label of the Relation instance. The explanation is based on a simple form where names of the people in question are placed. An example of an explanation generated is "Schjerfbeck, Helene and Swan, John Macallan had the same teacher Bastien-Lepage, Jules."

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX gvp: <http://vocab.getty.edu/ontology#>
PREFIX rel: <http://ldf.fi/schema/relations/>
CONSTRUCT {
  [] a rel:DRelation ;  # Relation instance to be constructed
    rel:relationSubject ?person ;
    rel:relationObject ?person2 ;
    rdfs:label ?description ;
    rel:relationType rel:sharedTeacher .
  }
WHERE {
  ?person a gvp:PersonConcept .
  ?person gvp:prefLabelGVP/gvp:term ?name .
  ?person2 a gvp:PersonConcept .
  ?person2 gvp:prefLabelGVP/gvp:term ?name2 .
  FILTER (?person != ?person2) # Related persons must be different
  ?person gvp:ulan1102_student_of ?teacher . # The persons have the same teacher
  ?person2 gvp:ulan1102_student_of ?teacher .
  ?teacher gvp:prefLabelGVP/gvp:term ?teacherName .
  BIND (CONCAT(CONCAT(CONCAT(CONCAT(CONCAT(?name, " and "), ?name2), # Explanation
    " had the same teacher "), ?teacherName), '.') AS ?description)
}
```

2 Faceted search user interface

To test the method, a web application was implemented using the Sampo-UI [7] framework. Relations were constructed from the ULAN KG and then ingested by a Fuseki[4] SPARQL server that the web application queries. The user interface (UI) of the demonstrator includes multiple facets on the left and the results in table format on the right, with human readable explanations of the connections and links to the related people. Faceted search means that the user can constrain the search incrementally, and the hit counts of each facet will update after each selection. Distribution of the hit counts of facets can be visualized for statistical analyses by pressing a pie chart button of a facet.

In our system the UI has separate facets for the subject and object of the connection and their properties. There are, for example, separate facets "Nationality A" and "Nationality B" that represent the nationality of the subject and the object respectively. A user can easily search, for example, for connections between two individual people, all connections for one person, connections between a certain person and persons of a certain nationality, and even connections between people of certain nationalities. For example in Fig. 1 the user has selected first "Shared teacher" as the connection type from the "Relation type" facet and "American" from the "Nationality A" facet. This limits the search to connections where an American artist shares a teacher with an other artist. The UI shows the

[4] https://jena.apache.org/documentation/fuseki2/

human readable explanations of each connection, and also the relative numbers of connections can be interesting. These can be immediately seen from the facets, but they can also be visualized by clicking a pie chart button in a facet. In this case the chart shows that American have most shared teacher connections with other Americans followed by French artists with Germans as a distant third.

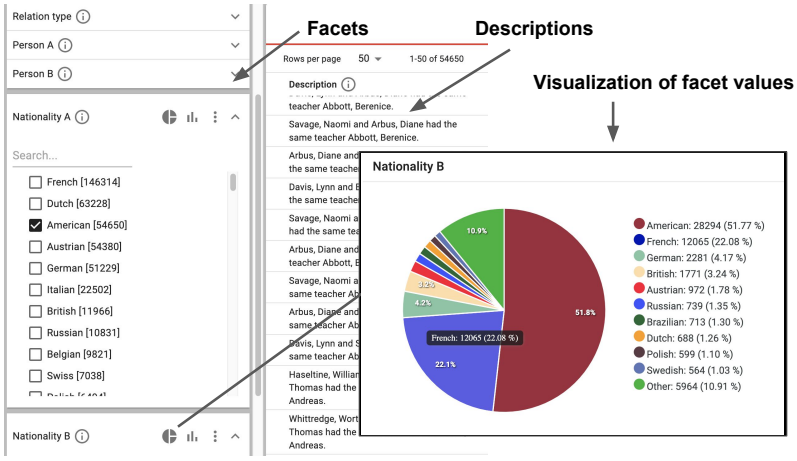

Fig. 1. An example of the UI showing connections involving American painters.

We treat all the Relation instances as directed connections, which means that, for example, every shared teacher connection is created twice, so that both artists of the connection are the starting point in one Relation instance. A benefit of the directed connections is that one can easily filter the connections using faceted search so that the subject and object of the connections have separate facets. It is then possible to easily search and visualize, for example, the comparative numbers of different nationalities of students of French painters, or any number of other combinations of individuals, groups, or relation types.

3 Related works and contributions

RS between artists in ULAN has been addressed as part of the CultureSampo[5] system [5,9], but with limited search functionality, and limited to teacher–student relations only. RelFinder and Explass [1] are examples of RS using faceted search, but in those applications faceted search is limited to only searching relations from a single entity that needs to be selected first. Our approach to RS is different from [10,8,4,3,11] by using domain specific cultural knowledge about connection

[5] http://www.kulttuurisampo.fi

types for filtering out "interesting" relations and for generating their natural language explanations.

ULAN KG includes rich data about connections between artists, and offers a good use case for our approach. We plan to publish a demonstrator application later in 2023. The faceted search of relations in ULAN KG could potentially be useful for art history research and art education, by showing novel connections in the data. Our aim is to use to use the method also to Wikidata and the international biographical KG under development in the InTaVia EU project[6] to discover more and different kind of international connections of artists.

References

1. Cheng, G., Zhang, Y., Qu, Y.: Explass: exploring associations between entities via top-k ontological patterns and facets. In: International Semantic Web Conference (ISWC). pp. 422–437. Springer (2014)
2. Cobb, J.: The journey to linked open data: The getty vocabularies. Journal of Library Metadata **15**(3-4), 142–156 (2015). https://doi.org/10.1080/19386389.2015.1103081, https://doi.org/10.1080/19386389.2015.1103081
3. Heim, P., Hellmann, S., Lehmann, J., Lohmann, S., Stegemann, T.: Relfinder: Revealing relationships in rdf knowledge bases. In: Proceedings of the 4th International Conference on Semantic and Digital Media Technologies (SAMT 2009). pp. 182–187. Springer (2009), http://dx.doi.org/10.1007/978-3-642-10543-2_21
4. Heim, P., Lohmann, S., Stegemann, T.: Interactive relationship discovery via the semantic web. In: Proceedings of the 7th Extended Semantic Web Conference (ESWC 2010). vol. 6088, pp. 303–317. Springer, Berlin/Heidelberg (2010), http://dx.doi.org/10.1007/978-3-642-13486-9_21
5. Hyvönen, E., Mäkelä, E., Kauppinen, T., et al.: CultureSampo – Finnish culture on the Semantic Web 2.0. Thematic perspectives for the end-user. In: Museums and the Web 2009, Proceedings. Archives and Museum Informatics, Toronto (2009)
6. Hyvönen, E., Rantala, H.: Knowledge-based relational search in cultural heritage linked data. Digital Scholarship in the Humanities (DSH) **36**, 155–164 (2021). https://doi.org/https://doi.org/10.1093/llc/fqab042
7. Ikkala, E., Hyvönen, E., Rantala, H., Koho, M.: Sampo-UI: A full stack JavaScript framework for developing semantic portal user interfaces. Semantic Web – Interoperability, Usability, Applicability **13**(1), 69–84 (2022)
8. Lohmann, S., Heim, P., Stegemann, T., Ziegler, J.: The RelFinder user interface: Interactive exploration of relationships between objects of interest. In: Proceedings of the 14th International Conference on Intelligent User Interfaces (IUI 2010). pp. 421–422. ACM (2010), http://doi.acm.org/10.1145/1719970.1720052
9. Mäkelä, E., Ruotsalo, T., Hyvönen: How to deal with massively heterogeneous cultural heritage data—lessons learned in CultureSampo. Semantic Web – Interoperability, Usability, Applicability **3**(1) (2012)
10. Sheth, A., Aleman-Meza, B., Arpinar, I.B., et al.: Semantic association identification and knowledge discovery for national security applications. Journal of Database Management on Database Technology **16**(1), 33–53 (2005)
11. Tartari, G., Hogan, A.: WiSP: Weighted shortest paths for RDF graphs. In: Proceedings of VOILA 2018. pp. 37–52. CEUR WS Proceedings, vol. 2187 (2018)

[6] https://intavia.eu/